



Detection and Localization of Texts from Natural Scene Images: A Hybrid Approach

Priyanka Muchhadiya
Post Graduate Fellow,
ECD, AITS,
Rajkot, Gujarat (India)

Abstract: Text detection in images or videos is an important step to achieve multimedia content retrieval. In this paper, an efficient algorithm which can automatically detect, localize and extract horizontally aligned text in images (and digital videos) with complex backgrounds is presented. The proposed approach is based on the application of a color reduction technique, a method for edge detection, and the localization of text regions using projection profile analyses and geometrical properties. The outputs of the algorithm are text boxes with a simplified background, ready to be fed into an OCR engine for subsequent character recognition. Our proposal is robust with respect to different font sizes, font colors, languages and background complexities. Text recognition and analysis includes many applications such as: license plate recognition, sign detection as well translation, helping tourists and blind persons to understanding environment, drawing attention of a driver, content-based image search and so on. Locating text in case of variation in style, color, as well as complex image background makes text reading from images more challenging. In this paper the various techniques available for detecting and recognizing text are explained, finally a hybrid approach using segmentation explained which can improve the qualitative texture analysis among other techniques.

Keywords: Text detection, Text localization, Text recognition.

I. INTRODUCTION

In a society driven by visual information and with the drastic expansion of low-priced cameras, vision techniques are more and more considered and text recognition is nowadays a fast changing field, which is included in a large spectrum, named text understanding. The classical OCR systems are not capable to recognize such text, as they are basically designed for bi-level text with a resolution of more than 150 dpi, typically obtained using at scanners. Several research works report on made to detect text regions from natural scene images, more robust and effective methods are expected to handle variations of scale, orientation, and clutter background. Algorithms for detection and recognition of text. However, these works are often directed towards the implementation of front-end image processing methods in order to enhance the image quality and pass it through classical OCRs. As indicative marks in natural scene images, text information provides brief and significant clues for many image based applications such as scene understanding, content-based image retrieval assistive navigation and automatic geo-coding. To extract text information from camera-captured document images (i.e., most part of the captured image contains well organized text with clean background), many algorithms and commercial optical character recognition (OCR) systems have been developed.

It used texture flow analysis to perform geometric rectification of the planar and curved documents performed topic-based partition of document image to distinguish text, white spaces and figures, Different from document images, in which text characters are normalized into elegant poses and proper resolutions, natural scene images embed text in arbitrary shapes, sizes, and orientations into complex background. It is impossible to recognize text in natural scene images directly because the off-the-shelf OCR software cannot handle complex background interferences and non-orienting text lines. Thus, we need to detect image regions containing text strings and their corresponding orientations. Although many research efforts have been are expected to handle variations of scale, orientation, and clutter background.

Texture analysis system consists of mainly four stages: text detection, text localization, text extraction, text recognition. We can use these stages text detection, localization, and extraction where text detection consists of determination of the occurrence of text in image, Text localization is the process of determining the location of text, shown in figure 1.

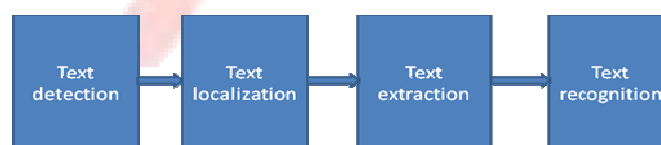


Figure-1 Basic block of text analysis

Usually extraction of different texts can be done using segmentation. So in text extraction stage the text components in images are segmented from background. After, the extracted text images can be converted into plain text using OCR technology. Now

with the help of text extraction and detection from natural images, which coupling of text-based searching technologies and optical character recognition (OCR), is now recognized as a key component which is present in the images.

But as we know that, text characters contained in images can be multicolor or gray-scale value with variable size, low resolution, and embedded in noisy backgrounds. Many experiments done on text recognition by applying conventional. OCR technology directly it leads to decrease rates of recognition. Therefore, for efficient detection and segmentation of text characters from the background is necessary to fill the gap between images and the text input of an OCR system. Previous methods classified into top-down methods and bottom-up methods. In Top down approach algorithms first detect text in images and then segment each of them into text and background. In bottom-up approach after segment images into regions and then group character regions into words. The recognition performance therefore relies on the segmentation.

As I explained earlier for detecting text from natural images have many methods. Here we consider only region based [1] component based method.

Now text detection and localization methods can be categorized into two parts: region-based and connected component (CC)-based both are based on segmentation finally will describe the method for text extraction and recognition.

A region-based segmentation method for text detection consists of following stages:

1. Text detection to estimate text existing confidence in local image regions by classification,
2. Text localization to cluster local text regions into text blocks,
3. Text verification to remove non-text regions for further processing.

And a connected component (CC) based segmentation method consists of following stage:

1. CC extraction to segment candidate text components from images;
2. CC analysis to filter out non-text components using heuristic rules or classifiers;
3. Post-processing to group text components into text blocks (e.g., words and lines).

These paper methods are categorized according to which technique will be used for detection and localization. That is CC based method and region based method. These methods are complementary to each other. If we merge this technique we get the robust output called hybrid approach. Remaining section arranges as follow II Techniques/Method for text detection and localization III Proposed Work IV Simulation Results V Conclusion VI Future Scope.

II. TEXT DETECTION METHODS

For detecting text from natural images have many method. Here we consider only region based and connected component based method.

1. A region-based method:

It consists mainly following stages:

- Text detection to estimate text existing confidence in local image regions by classification
- Text localization to cluster local text regions into text blocks
- Text verification to remove non-text regions for further processing.

This new approach to accurately detect text in color images possibly with a complex background. This algorithm is based on the combination of connected component and texture feature analysis of unknown text region contours. First, describe color image edge detection algorithm to extract all possible text edge pixels. This algorithm is mainly based on following properties of text

1. Contour gradient
2. Structural information
3. Texture property

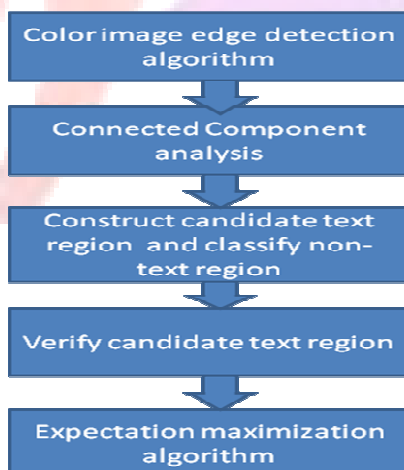


Figure-2 Region based algorithm

2. A connected component based method:

It consists of following stages:

- CC extraction to segment candidate text components from images
- CC analysis to filter out non-text components using heuristic rules or classifiers
- Post-processing to group text components into text blocks (e.g., words and lines)

This algorithm has following stages to detect and localized text from natural scenes images.

Text Localization:

Text recognition is generally divided into four steps: detection, localization, extraction, and recognition. The detection step roughly classifies text regions and non-text regions.

Text Region Localization

In this approach, the edges are considered a very important portion of the perceptual information content in an image. Mathematical Morphology is a topological and geometrical based approach for image analysis.

Let $S_{m,n}$, denote a structure element with the size $m \times n$, where m and n are odds and larger than zero and $I_{x,y}$, denote a gray-level input image. According to the definition of $S_{m,n}$, the smoothing, dilation, erosion, closing, opening, and other operations are mathematically represented as:

Smoothing operation:

$$E_{S_{m,n}}(I(x,y)) = \frac{1}{mn} \sum_{i=-m/2}^{m/2} I(x+i,y+j) S_{m,n}(i,j) \dots\dots\dots 1$$

Dilation operation:

$$I(x,y) \oplus S_{m,n} = \max_{|i| \leq m/2, |j| \leq n/2} I(x-i,y-j) S_{m,n}(i,j) \dots\dots\dots 2$$

Erosion operation:

$$I(x,y) \ominus S_{m,n} = \max_{|i| \leq m/2, |j| \leq n/2} I(x-i,y-j) S_{m,n}(i,j) \dots\dots\dots 3$$

Closing operation:

$$I(x,y) \bullet S_{m,n} = (I(x,y) \oplus S_{m,n}) \ominus S_{m,n} \dots\dots\dots 4$$

Opening operation:

$$I(x,y) \circ S_{m,n} = (I(x,y) \ominus S_{m,n}) \oplus S_{m,n} \dots\dots\dots 5$$

Using the above operations, boundaries are detected. Then the algorithm is tried to find out the connected components in a binary image using Connected Component Labeling method. The threshold value is obtained using the function given below:

Thresholding operation:

$$T(I(x,y)) = \begin{cases} 255, & \text{if } I(x,y) > T \\ 0, & \text{otherwise} \end{cases} \dots\dots\dots 6$$

From the threshold image, a connected component algorithm is performed and to find the labeled regions of connected pixels which have the same value.

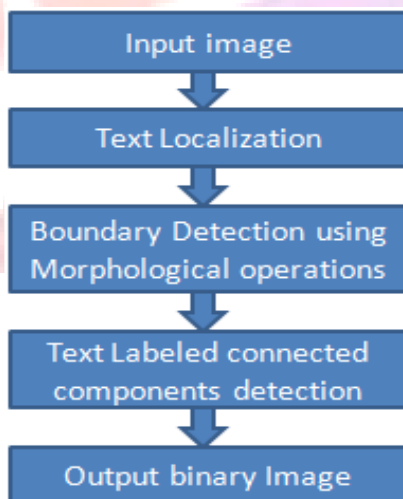


Figure-3 Connected component based approach

III. PROPOSED WORK

In region-based methods, there are some drawback like speed of operation is comparatively slow; the performance is sensitive to text alignment orientation. For CC-based methods 1) cannot segment text components accurately without prior knowledge of text position and scale 2) designing fast and reliable connected component analysis is difficult since there are many non-text components which are easily confused with texts when analyzed individually. To overcome these difficulties from region based and connected component based methods I have developed a new approach known as hybrid approach. This has complementary advantages and disadvantages of both methods for text detection and recognition for obtaining better output.

So the proposed work is hybrid approach [1]. In that I have used both methods region based segmentation and connected component based segmentation.

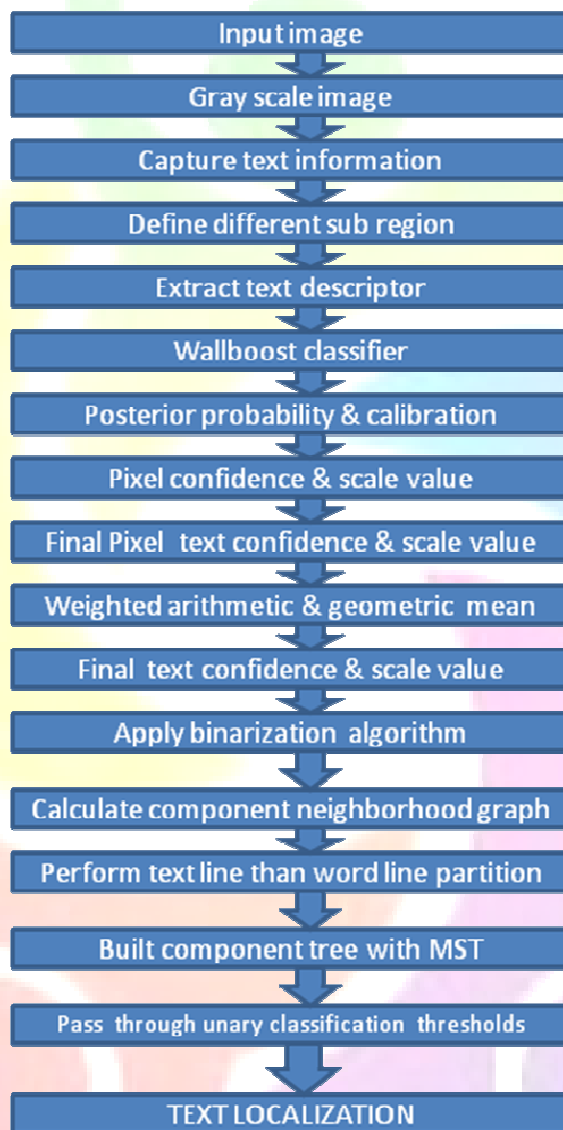


Figure-4 Hybrid approach

Description:

Here I give input image which converted to gray scale then apply tree region detector to get the text confidence using Wald boost classifier. Then using posterior probability calculates the text confidence and scale map. Apply text confidence and scale map to get the final text confidence and scale map. Using scale map image segmentation will be performing. For labeling the component with CRF model we use text confidence. Then apply the minimum spanning tree for clustering component. Then apply line/word partition to get the localization of text.

Performance Parameters:

Mainly three performance parameters are used to report results practically recall, precision, and false alarm rate, in which recall and precision rate must be 100% ideally or as high as possible and false alarm should 0% or low as possible. These all parameters are defined as:

$$\text{Recall} = \frac{\text{correct detects}}{\text{correct detects} + \text{missed detects}} \dots\dots\dots 1$$

$$\text{Precision} = \frac{\text{correct detects}}{\text{correct detects} + \text{false alarms}} \dots\dots\dots 2$$

$$\text{False Alarm Rate} = \frac{\text{False Alarms}}{\text{Correct Detections}} \dots\dots\dots 3$$

$$f - \text{metric} = \frac{1}{\frac{0.5}{\text{precision}} + \frac{0.5}{\text{recall}}} \dots\dots\dots 4$$

IV. SIMULATION RESULTS

Hereby the implementation results for proposed approach is there in which different kind of text image with blurred, complex text as well as complex background, also with different alignment of text image are simulated using MATLAB tool.

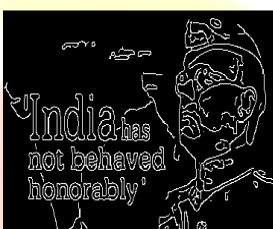
The result shows the detection text successfully with some enhanced result.



Original Image



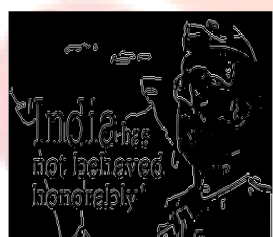
Gray scale image



Edge of an image



Text confidence region



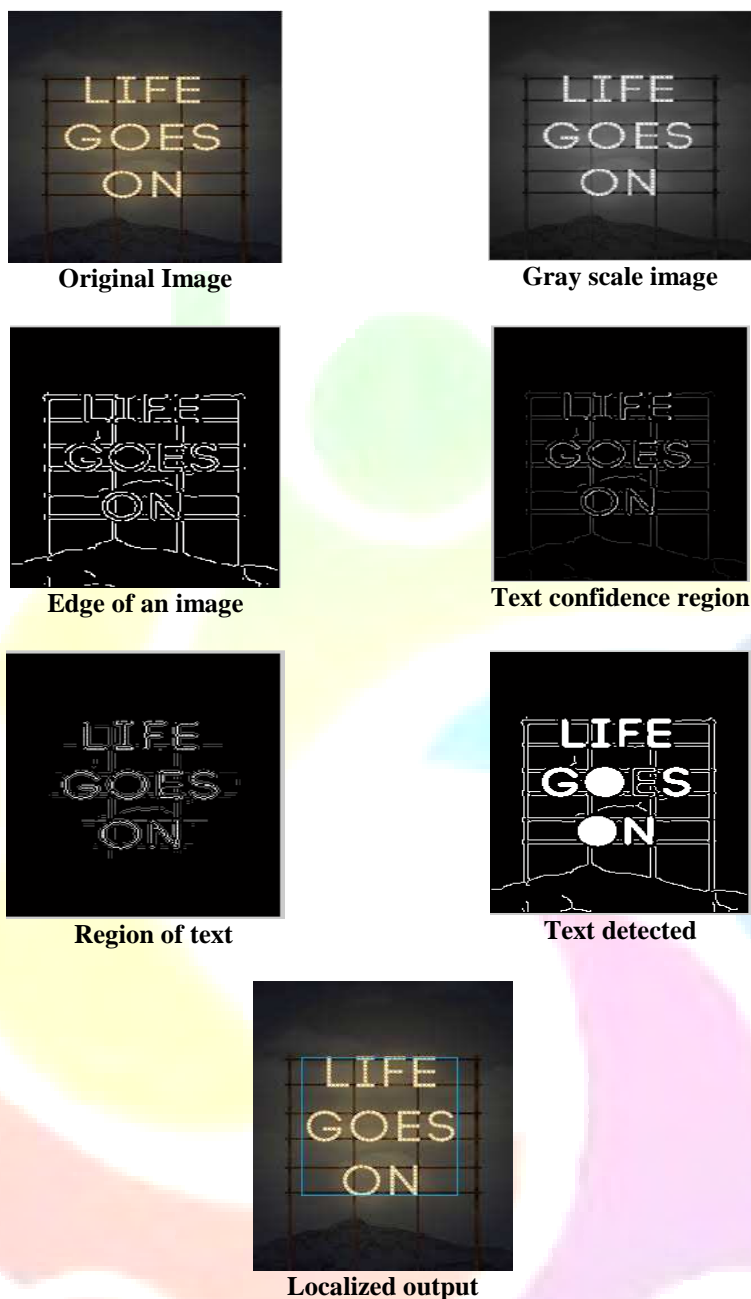
Region of text



Text detected



Localized output



V. CONCLUSION

From this summary we clarify that the best method for text detection and localization is the hybrid or proposed approach which over comes most of the disadvantages of present algorithms. In existing algorithm region based text detection techniques the text will be detected and localize be texture analysis by finding region, Due to that detection and localization accurately in presence of noise as well. In Connected Component-based algorithm directly segment candidate text components by edge detection Having lower computation cost and the located text components can be directly used for recognition. In region-based algorithm speed is relatively slow, and the performance is sensitive to alignment and orientation of text. For Connected Component -based algorithm we cannot segment text components accurately without prior knowledge of text position and scale and designing fast and reliable connected component analyzer is difficult since there are many non-text components which are easily confused with texts when analyzed individually. so in the proposed hybrid algorithm these two methods are get merged to get the better output. From above experimental result it is clear that we can achieve that.

VI. FUTURE SCOPE

The proposed hybrid approach also having some drawbacks like method fails on some hard-to-segment texts and Speed still need to be increased. So the PG scholar may work to overcome these drawbacks as a future work.



REFERENCES

1. Yi-Feng Pan, Xinwen Hou, and Cheng-Lin Liu, "A Hybrid Approach to Detect and Localize Texts in Natural Scene Images", IEEE Trans on Image Processing, VOL. 20, NO. 3, March 2011
2. Y. X. Liu, S. Goto, and T. Ikenaga, "A contour-based robust algorithm for text detection in color images", IEICE Trans. Inf. Syst., vol. E89-D, no. 3, pp. 1221-1230, 2006
3. X. L. Chen, J. Yang, J. Zhang, and A. Waibel, "Automatic detection and recognition of signs from natural scenes," IEEE Trans. Image Process., vol. 13, no. 1, pp. 87-99, Jan. 2004 [4] M. Cai, J. Song, and M. R. Lyu, "A new approach for video text detection," in Proc. Int. Conf. Image Process., Rochester, NY, Sep. 2002, pp.117-120C.
4. Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell., 23(11):1222-1239, 2001
5. Liang J., Doermann D., and Li H., "Camera-based analysis of text and documents: A survey," Int. J. document Anal. Recogn., 2005; 7: 84-104..
6. Wu V., Manmatha R., and Riseman E. M., "Text Finder: An Automatic System to Detect and Recognize Text in Images," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009; 21: 1224-1229.
7. C. Yi and Y. Tian, "Text string detection from natural scenes by structure-based partition and grouping", IJFST vol.19, no. 12, 2011.
8. Elie Bursztein, Matthieu Artin, Stanford University,
9. Text-based CAPTCHA Strengths and Weaknesses ACM(CSS 2011). Pranob, K Charles, V. Harish, "A Review on the Various Techniques used for Optical Character Recognition" IJST Vol. 2, Issue 1, Jan-Feb 2012.
10. Miss. Poonam B. Kadam, Mrs. Latika R. Desai "A Hybrid Approach to Detect and Recognize Texts in Images" IJRCCT vol 2 issue 7, 2013.